

## 1 はじめに

一局のゲームは、個々の着手の着手位置を記録した棋譜によって記述され、棋譜からは、互いの石の配置などの静的局面情報や着手系列によって局面がどのように変化していったかなど、ゲームの進行に関する情報を全て再現することができる。また、囲碁は別名「手談」とも言われるように、その対局は着手を通じたコミュニケーションとみなすことができ、個々の着手にはこれ以外にもその局面における役割やプレイヤーの意図などの様々な情報が内在されていると考えられる。このような情報を棋譜から抽出することは、ゲームの内容を理解するためには欠かせない作業であり、また、結果として得られたこれらの情報は、局面認識や着手決定のためのパターン知識など幅広い利用が期待できる。

筆者らは、この問題に対して、自然言語処理とのアナロジーから、棋譜を個々の着手を符号化してできたテキスト(棋譜テキスト)であるとみなし、この棋譜テキストに対して適切な言語モデルを作成し言語処理技術を適用するというアプローチを行なっている [1][2][3]。そして、棋譜テキストに対する言語モデルが、盤面認識や着手候補選出などゲームプレイシステムの中核として利用できるだけでなく、ゲーム記述言語と自然言語との相互変換を通じて、棋譜からの解説文生成や自然言語による棋譜データベース検索など、知的ゲームに関する自然言語システムへと幅広く応用することができ、ひいては、言語以外の分野における人間の言語的思考過程の解明へと繋がることを期待している。

特に囲碁は  $19 \times 19$  の広大な盤面と基本的に空点にはどこにでも打てるという着手の自由度のため、力技の探索だけでは到底太刀打ちできない。また、個々の石には区別がなく、盤上での他の石との位置関係から役割が決まるため局面評価が難しい。これらの特質は、チェスにかわる新たな AI のテストベッドとして囲碁が注目される要因となっている [4]。囲碁研究には、認知科学、機械学習、組合せ数学、さらには自然言語処理など様々な分野の研究が関連し、個々の技術の進歩、そしてその集大成としてのコンピュータ囲碁の進歩が、今後の AI 研究の牽引車となることを期待している。

## 2 定形手順の獲得法

一般に定石とは序盤に部分的に出現する一定の石の形およびそこに至る手順を指すが、中盤の定石という類のものもあることから分かるように、序盤に限らず、碁の法則から導かれる一定の理にかなった着手の応酬というものが存在する。定型手順は、過去の棋譜中のさまざまな場所に頻繁に出現しており手順の長短、局面範囲にかかわらず、一局の棋譜を通じて定型であると認められる手順をすべて獲得することが望まれる。囲碁におけるこのような定型手順は、自然言語テキストにおける定型表現と類似した特徴を持っている。

NL テキストの定型表現	定型手順 (定石)
頻繁に使用される	頻出
ひとまとまりの表現	単位性のある連続着手
単語よりも大きい単位	ある程度の長手順

日本語のように単語の間に空白を置かずにはべた書きされる

テキストデータの場合、単語やそれが連なった定型的な表現を抽出することを目的として、辞書と文法による形態素解析を行わずに、文字列の出現頻度情報のみを用いて定型表現を抽出する様々な手法の研究が行なわれており、これらの手法は棋譜テキストからの定型手順獲得に応用できる。

### 2.1 着手の符号化

$n$ -gram ではパターンの一貫性は文字列の一貫性によって判断される。したがって、手順パターンの定型性評価が正しく行なわれるためには、着手の符号化において以下のことが要請される。

1. 個々の着手に対して、時間的、空間的に局所的な情報のみを符号化する。
2. 盤上での回転、鏡像、移動の関係にある手順が同一の符号列になる。
3. 形の異なる手順は、同一の符号列にならない。

これらの要請 (1)–(3) を満足する符号化法として以下に示す符号化を行なう。

**相対符号化法 ( $E_R$ )** 直前の相手方の着点を座標原点とする。座標軸 ( $x$  軸,  $y$  軸) の選択と正負の方向は、直前の着点が図 1 で示した 8 つの領域のどれに属するかに応じて決定する。そして得られた座標軸を用いて現在の着点の相対位置  $(x, y)$  を求め、符号  $c_{x,y}$  を与える。

例えば、現在の着点が  $(3, 9)$ 、直前の着点が  $(4, 7)$  の場合には、符号  $c_{2,-1}$  を与える。

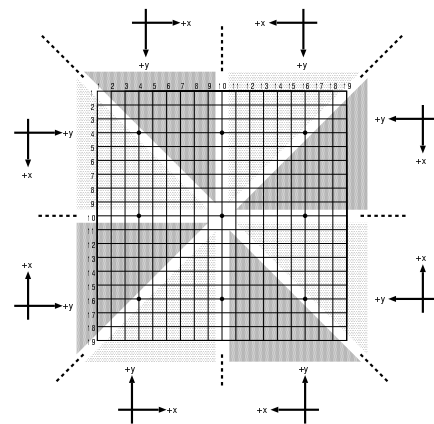


図 1: 直前の着点と座標軸との対応

この符号化法を用いて、実際に棋譜の符号化を行なった。棋譜データベースとしては、日本棋院「棋譜データ集 96」CD-ROM に収録されているプロ棋士の対局約 34,000 局、総手数約 700 万手分のデータを用いた。その結果、アルファベット総数は 738、頻度 2 以上の部分文字列の異なり数は  $2.25 \times 10^7$  となった。次に、同一符号に符号化された手順列中に形の異なる手順がどの程度含まれているかを符号長毎に集計した結果を表 1 に示す。これを見ると、 $E_R$  符号化によって要請 (3) がかなりの程度満足されていることがわかる。

符号列長	異形数		符号列長	異形数	
	最大	平均		最大	平均
4	6	1.30	7	4	1.01
5	5	1.13	8	3	1.00
6	4	1.04	9以上	2	1.00

表 1:  $E_R$  符号化における同一符号列の異形数

## 2.2 出現頻度に基づく定型性評価

自然言語テキストからの定形表現の獲得において、複数の  $n$  に対する  $n$ -gram を用いて表現単位の定型性を評価する様々な試みが行なわれている。正規化頻度法 [5] は、 $n$ -gram を直接用いる方法に比べて断片的な文字列の抽出を避けるよう改良がなされているが、基本的にセグメンテーションを行わない手法であるため、それでもなお互いに重なりを持つ文字列や断片的な文字列を抽出してしまうことがある。一方、隣接文字エントロピー法 [6] は、単位性の認定を重視して定型手順を獲得する手法であるが、一般に出現頻度の高い短い手順を優先して抽出する傾向にある。そこで我々は、 $n$ -gram を用いて文字列から定形表現を直接切り出すための新しい手法として、部分列頻度プロファイル法 (Substring Frequency Profiling Method) を用いて定型性の評価を行なう。

### 2.2.1 部分列頻度プロファイル法 (SFP 法)

SFP とは、注目する部分列の長さ (窓幅)  $w$  を固定し、ある棋譜データについて  $i-w+1$  手目から  $i$  手目までの長さ  $w$  の着手列がデータベース中に出現した頻度を各  $i$  に対して記録したものである。

複数の  $n$  に対する  $n$ -gram を用いた場合、ある文字列  $x$  が対象テキスト中出现する頻度  $f(x)$  と  $x$  の部分文字列  $y$  の出現頻度  $f(y)$  の間には  $f(x) \leq f(y)$  が成立する。すなわち、ある文字列の部分列は元の文字列よりも出現頻度が高いため、注目している部分列が頻出する基本定石手順の内部にあるときはその出現頻度は高い値をとる。一方、定石は単位性のある連続着手であるため、定石が一段落した境界をまたぐ部分やその外部では出現頻度は低い値となる。したがって、SFP の値は一般に図 2 に示すような形状をなす。そこで、SFP の値の変化量に適切な閾値を設定して“山”を切り出すことによって高頻度かつ単位性のある手順パターンが獲得できる。

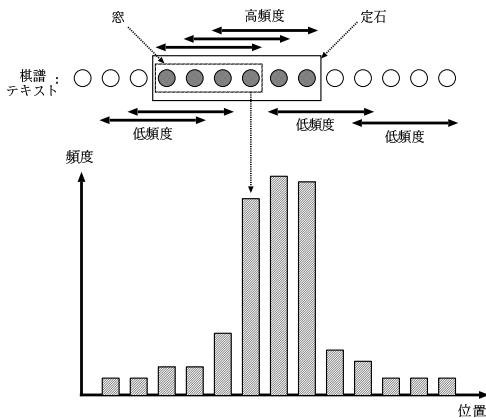


図 2: 部分列頻度プロファイル (SFP) 法

図 3 の棋譜に対して  $w = 6$  で作成した SFP を図 4 に示す。図 3 では、序盤に右下と左上に基本定石が出現しており、図 4 の SFP に基本定石部分が明瞭に浮き出ていることがわかる。また、これ以外にも 6 つの“山”が認められ、これに対応する部分を抽出したものを図 5 に示す。

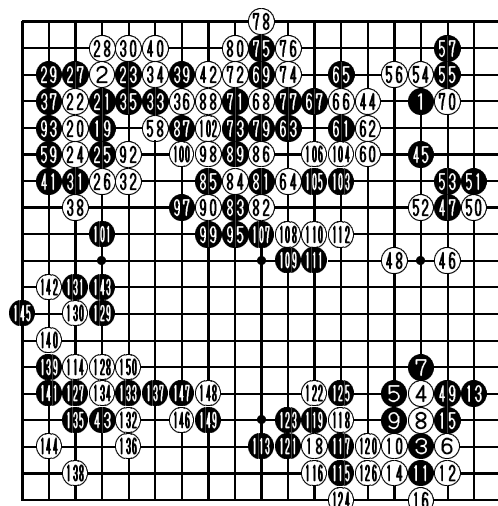


図 3: サンプル棋譜 (「棋譜データ集 96」より)

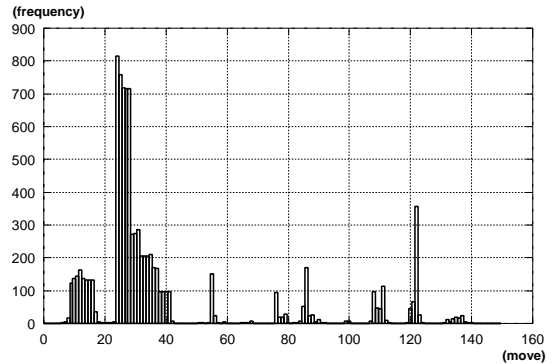


図 4: 窓幅 6 の SFP

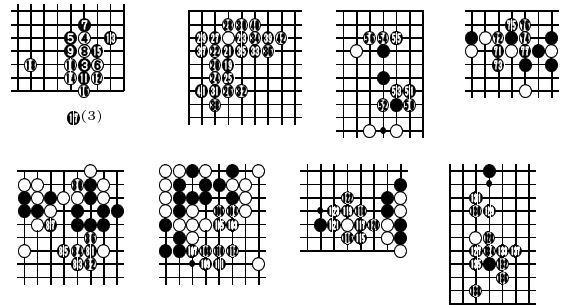


図 5: 獲得された定形手順

## 3 おわりに

棋譜から定石などの定形的な手順を獲得するために、棋譜を着手が符号化されてきたテキストであるととらえ、 $n$ -gram に基づいて定型性を評価する手法を示した。

## 参考文献

- [1] 梶山, 中村: “囲碁の着手記号列に対する確率文法モデルの作成”, GPW'99, pp.161-168, (1999).
- [2] 中村, 梶山: “棋譜テキストからの特徴素抽出とそれを用いた棋譜の分類”, 情処 GI 研究会 GI 2-11, pp.79-86, (2000).
- [3] 中村: “着手記号列の出現頻度に基づく囲碁棋譜からの定型手順獲得”, 情処論文誌, Vol.43, No.10, pp.3030-3039, (2002).
- [4] 佐々木, 清, 中村: “ゲーム情報学研究の事例: 囲碁”, 情報処理, Vol.44, No.9, pp.910-915, (2003).
- [5] 中渡瀬: “統計的手法によるテキストからのキーワード抽出法”, 信学会データ工学研究会 DE 95-2, pp.9-16, (1995).
- [6] 下畑他: “隣接文字の分散値を用いた定型表現の自動抽出”, 情処 NL 研究会 NL 110-11, pp.71-78, (1995).